

# Final Exam

EC 204: Empirical Economics 2

Due on: Thursday August 13, 2020 by 10:00AM (EST)

## Instructions

The objective of this assessment is to test your knowledge of what we have covered over the last six weeks. You have 48 hours to complete this exam. You may refer to the course material when completing this assessment. **You cannot discuss this assessment with any individual(s). Doing so will be considered cheating.**

You are responsible for making sure that you understand each question clearly. If not, be sure to ask me any questions that might arise. If you use any STATA commands/options/statistical methods that have not been discussed in class, briefly describe what they are and justify why you chose to use them.

All relevant data files are provided in the “Final” folder (in .dta format as always). **Be sure to submit your .do file along with your responses to the exam questions. Everything must be compiled in one file.**

Good luck!

## Question 1

The data set `HPRICE.DTA` contains a set of variables for each of 506 neighborhoods. The variables in the data set are defined below:

- `nnumber` – neighborhood number
  - `price` – median neighborhood house price (in dollars)
  - `crime` – crimes committed in neighborhood per capita
  - `nox` – atmospheric concentration of nitrous oxides (p/100m)
  - `rooms` – average number of rooms in neighborhood houses
  - `dist` – weighted distance of neighborhood to 5 employment centers
  - `radial` – index of neighborhood access to radial highways
  - `proptax` – property tax per \$1000 valuation in neighborhood
  - `stratio` – student-teacher ratio in neighborhood schools
- (a) What determines the median house price in these neighborhoods? Use regression analysis to estimate a model of house prices. Explain and defend your choice of specification and estimation method (independent variables chosen, functional form, standard errors, etc.). Report your results.
- (b) Suppose that you have been hired by the board of directors of neighborhood number 130's homeowner's association. The board believes that their neighborhood's property is undervalued and wants to increase house values. They have asked you to advise them on the following propositions using your statistical analysis. For each, write a short response supported by your estimates and/or calculations
- (i) Are board members justified in believing that property values for neighborhood 130 are too low, given the neighborhood's characteristics?
  - (ii) Joe, one of the board members, has argued that improving a major road connecting the neighborhood to downtown would increase property values. The road improvement would increase the value of the radial access index from 4 to 8, but the increased traffic would increase nitrous oxide concentrations from 6.24 to 6.50. Does your analysis support Mr. Commuter's argument?
  - (iii) Board member Sarah is worried about neighborhood crime. She has proposed an increase in police service that would require an increase in property taxes from 43.7 to 47.7 and lower crime from 0.881 to 0.6. Would this increase property values?

## Question 2

From a sample of  $N = 45$  firms, we observe the amount of scrap $_i$  (waste from production) as well as hrsemp $_i$  (the number of hours of job training per employee) in each firm.

- (a) We first run an OLS regression of  $\log(\text{scrap}_i)$  on  $\text{hrsemp}_i$

$$\log(\text{scrap}) = \beta_0 + \beta_1 \text{hrsemp}_i + u_i$$

Given the specification of the regression, what is the interpretation of the coefficient  $\beta_1$ ?

- (b) We estimate  $\hat{\beta}_1 = -0.0076$  with standard error of 0.0045. Discuss the sign, statistical significance of the estimate, and its magnitude (consider, for instance, increasing  $\text{hrsemp}_i$  from 0 hours, the minimum in the sample, to 20 hours, the average in the sample).
- (c) Is the exogeneity assumption,  $\text{Cov}(u_i, \text{hrsemp}_i) = 0$  credible here? Discuss why or why not. Someone suggests using  $\text{grant}_i$  as an instrumental variable: a variable indicating whether a firm received a job training grant. How should the grant be awarded for this to be a valid instrument?
- (d) We estimate the instrumental variable regression using  $\text{grant}_i$  as an instrument, we find:

$$\begin{aligned} \widehat{\log(\text{scrap})}_i &= \underset{(0.127)}{-0.033} - \underset{(0.008)}{0.014} \widehat{\text{hrsemp}}_i \\ \widehat{\text{hrsemp}}_i &= \underset{(1.56)}{0.51} + \underset{(3.13)}{27.88} \text{grant}_i \end{aligned}$$

Discuss the sign, statistical significance of the estimate, and its magnitude (consider, for instance, increasing  $\text{hrsemp}_i$  from 0 hours, the minimum in the sample, to 20 hours, the average in the sample). Compare your answers with the OLS estimates.

- (e) Is the instrument weak? Discuss.
- (f) Given the data at hand, is it possible to test if the instrument is exogenous?

### Question 3

An important relationship in macroeconomics is that between inflation and unemployment (the Phillips Curve). The two variables are traditionally related as follows:

$$\pi_t - \pi_t^e = \beta_1(u_t - u^*)$$

where  $\pi_t$  refers to inflation at time  $t$ ,  $\pi_t^e$  refers to expectation of inflation at time  $t$ , and  $u^*$  is the natural rate of unemployment. One way to proxy for inflation expectations is by using the average expected rate of inflation according to professional forecasters. For this question, you will make use of the dataset `SPF_PC.DTA`. The dataset includes the following variables:

`date` – Date variable (requires no further manipulation)  
`inf_rev` – Revised inflation  
`inf_rt` – Real-time inflation  
`infexply` – Average one-year ahead inflation expectation  
`urate` – Unemployment rate

Two notes: (1) I have constructed `infexply` so that you can directly use it as a proxy for  $\pi_t^e$  (no need to deal with matching horizon of forecast with period in which inflation was realized). (2) You can ignore serial correlation and just use the `regress` command for this entire question.

- (a) Estimate the Phillips Curve relationship described above. Use `inf_rt` as your measure of inflation. Report and interpret your results. How much does a one point increase in unemployment affect unanticipated inflation?
- (b) From your point estimates, determine the implied natural rate of unemployment,  $u^*$ .
- (c) Now estimate the regression in (a) but only for the years 2000 onward. Do your results change? What does this mean for the relationship between inflation and unemployment?
- (d) Over time, the government revises macroeconomic data, sometimes to address measurement error issues. Re-estimate the Phillips Curve with the full sample using `inf_rev` rather than `inf_rt`. How do your results compare to those in (a)?
- (e) Given your results in (d), determine the implied natural rate of unemployment,  $u^*$ . Compare with your answer for (b).

## Question 4

A paper by Gregory Martin and Ali Yurukoglu (2017) studies the effects of exposure to cable news bias on voting patterns. In particular, they wish to run the following regression

$$\text{VoteShare}_z^R = \beta_0 + \beta_1 \text{MinutesWatched}_z + \varepsilon_z \quad (1)$$

where the dependent variable is the Republican party's vote share in zip code  $z$  during the 2008 presidential election and the independent variable is the average number of minutes of Fox News watched (Fox News is understood to be a conservative-leaning cable news network). The error term is  $\varepsilon_z$ .

- (a) In words (no math) state the zero conditional mean assumption which must hold in order for  $\beta_1$  to have a causal interpretation. Why is there reason to believe that the assumption fails in this case?
- (b) The authors propose an instrument for exposure to media bias: the channel positions of news channels in cable television lineups. The channel position is the ordinal position of news channels in the cable lineup, which varies by zip code. What are two conditions that must hold in order for this to be deemed a "good" instrument? Which of these conditions can be tested and which cannot?
- (c) Consider the condition in (b) that you determined cannot explicitly be tested. Do you believe that this condition holds? Explain why or why not.
- (d) The table below details the results of the first stage regression (with standard errors reported below the point estimates, in parenthesis). Interpret the sign of the first stage regression coefficient and discuss whether the instrument is weak.

Fox News Channel Minutes per Week	
Fox News Channel position	-0.174 (0.028)
Observations	59,541
$R^2$	0.377

Table 1: First Stage Regression

- (e) Suppose that you add an exogenous variable to regression (1) denoting the share of women in zip code  $z$ , ( $\text{FemaleShare}_z$ ). Write out the reduced form equation for this regression model.

## Question 5

In a paper by Joel David, Espen Henriksen and Ina Simovoska titled “The Risky Capital of Emerging Markets,” the authors present two new facts about the return to capital in emerging markets. In particular they note that:

- (i) emerging markets exhibit higher average returns to capital
- (ii) emerging markets are more highly exposed to US capital returns

Make use of the dataset `RISKYCAP.DTA` for this question. This data set includes information on returns to capital (`return`) across countries over several years. There is also a variable specific to the return to capital in the US (`return_us`). In addition, there is a variable called `portfolio` which classifies countries based on their real GDP in 2008. Portfolio 1 consists of those countries with the lowest real GDP whereas Portfolio 3 consists of the wealthiest countries. Lastly, there are three dummies for the `portfolio` variable (i.e. a dummy that equals 1 when portfolio equals 1, etc.)

- (a) Verify empirical fact (i). To do so, run a regression of `returns` on the different portfolio dummies. What is your base group? Interpret your results and verify that emerging markets indeed exhibit higher returns.
- (b) A way to measure the “riskiness” of an asset is to determine its “beta”. Here we will verify empirical fact (ii). First, regress country returns on the US return for the poorest set of countries. Then repeat for the wealthiest countries. Report and compare the two sets of estimates (no formal comparison tests required). How do these results provide evidence in favor of empirical fact (ii)?
- (c) Now verify these two facts within the *same* regression. Build on the regressions you specified in (a) and (b) by incorporating potential interactions judiciously. (Note: I just want to see a single regression that can speak to both the higher average returns and the greater exposure to US returns among emerging markets.)